UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/057,331	01/24/2002	Andrei Z. Broder	5598/153US	1535

7590 10/18/2007

Seth Ostrow
BROWN RAYSMAN MILLSTEIN FELDER AND STEINER LLP
900 Third Avenue
New York, NY 10022-4728

EXAMINER	
HILLERY, NATHAN	

ART UNIT	PAPER NUMBER
2176	

MAIL DATE	DELIVERY MODE
10/18/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Application Number: 10/057,331
Filing Date: January 24, 2002
Appellant(s): BRODER ET AL.

Timothy Bechen
For Appellant

EXAMINER'S ANSWER

This is in response to the appeal brief filed 7/11/07 appealing from the Office action mailed 7/11/06.

(1) Real Party in Interest

A statement identifying by name the real party in interest is contained in the brief.

(2) Related Appeals and Interferences

The examiner is not aware of any related appeals, interferences, or judicial proceedings which will directly affect or be directly affected by or have a bearing on the Board's decision in the pending appeal.

(3) Status of Claims

The statement of the status of claims contained in the brief is correct.

(4) Status of Amendments After Final

No amendment after final has been filed.

(5) Summary of Claimed Subject Matter

The summary of claimed subject matter contained in the brief is correct.

(6) Grounds of Rejection to be Reviewed on Appeal

The appellant's statement of the grounds of rejection to be reviewed on appeal is substantially correct. The changes are as follows: the rejection of claims 1, 2, 5 – 12 and 14 – 16 under 35 USC 103(a).

Grounds of Rejection WITHDRAWN

The following grounds of rejection are not presented for review on appeal because they have been withdrawn by the examiner. The rejection of claims 3 and 4 under 35 USC 103(a) are withdrawn.

(7) Claims Appendix

The copy of the appealed claims contained in the Appendix to the brief is correct.

(8) Evidence Relied Upon

Google Definitions

6112203 Bharat et al. 8-2000

(9) Grounds of Rejection

The following ground(s) of rejection are applicable to the appealed claims:

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 1, 2, 5 – 12 and 14 – 16 are rejected under 35 U.S.C. 103(a) as being unpatentable over Bharat et al. (US 6112203 A).

Regarding independent claim 1, Bharat et al. teach that *the set of documents can be produced by combining the set of results from a Web search engine in response to a user query (which we call the `start-set`), with pages that either link to or are linked from the start-set documents* (Column 3, lines 3 – 15), which meet the limitation of **receiving a document to be processed; locating a set of documents that include hyperlinks to the document**.

Bharat et al. also teach that *a simple approach uses the relevance weights of all of the nodes to decide whether or not to eliminate a page for user consideration. For example, prune all nodes whose relevance weight is below a predetermined threshold. The threshold can be picked in a number of ways* (Column 6, lines 22 – 27), which meet

the limitation of **for each token: determining a weight for the token, determining whether the weight assigned to the token exceeds a threshold token weight**.

Bharat et al. do teach that *in order to help users locate Web pages of interest, a search engine 140 maintains an index 141 of Web pages in a memory, for example, disk storage* (Column 4, lines 9 – 11) and that *we provide an improved ranking method 200 that can be implemented as part of the search engine 140. Alternatively, the method 200 can be implemented by one of the clients 110, or some other computer system on the path between the search engine and the clients* (Column 4, lines 23 – 27), which meet the limitation of **indexing the document under the token, if the token weight assigned to the token exceeds the threshold token weight**.

Bharat et al. do not explicitly teach **retrieving anchortext associated with at least one of the hyperlinks**, and **parsing the anchortext into one or more tokens**.

Bharat et al. teach that *the nodes in the start set are first scored according to their connectivity, and the number of terms of the query that appear as unique sub-strings in the URL of the represented documents. The score is a weighted sum of the number of directed edges to and from a node and the number of unique sub-strings of the URL that match a query term* (Column 3, lines 3 – 15).

It should be noted that anchortext could be the string of the URL making the sub-strings of the URL parsed tokens.

Thus, it would have been obvious to a person of ordinary skill in the art to try parsing anchortext that can be a string other than the string of the URL in an attempt to provide an improved ranking system, as a person with ordinary skill has good reason to

pursue the known options within his or her technical grasp. In turn, because anchortext as claimed has properties, simply a label or string, predicted by the prior art, it would have been obvious to parse any type of anchortext.

Regarding dependent claims 2, 5, 10 and 11, Bharat et al. teach that *specifically, in step 220, we score each page p of the input set 201 to determine a value Score(p) 225. Let n_p be the node representing page p. The score is determined by: Score(p)= in_degree + 2 x (num_query_matches) + out_degree, where in_degree is the number of edges pointing at node n_p, num_query_matches is the number of unique sub-strings of the URL of the page p that exactly match a term in the user's query* (Column 5, lines 57 – 64), which meet the limitation of **including in the index an indication of weight for each token under which each page is indexed**, and that **the weight of each token is based on its frequency of occurrence within the index**.

Regarding dependent claim 12, Bharat et al. teach that *next, we assign a relevance weight to a subset of the nodes 212. The relevance weight measures the similarity between the represented page and the query topic. As stated above, the topic implied by the user is probably broader than the query itself. Thus, matching the words of the query with the page is usually not sufficient. Instead, as described in detail below, we use a subset of the pages of the start set 201 to define a broader query topic "Q", and match the pages "P" represented in the graph with the broader query topic to determine the relevance weights of the nodes 212. Our invention is motivated by the*

observation that not all pages represented by nodes in the n-graph 211 are equally influential in deciding the outcome of our ranking process (Column 5, lines 21 – 33). Bharat et al. do not explicitly teach assigning the token to the beginning of the page; however, one of ordinary skill in the art at the time of the invention would be motivated to alter the invention of Bharat et al. to meet the limitation of **assigning to the token a location within the index that corresponds to the beginning of the page being indexed**, since the skilled artisan is well aware that by default when a webpage is returned to a user, the beginning of the webpage is returned; thus, the skilled artisan would want to point the user to the beginning of the page so that the default behavior, to which most users are accustomed, is mimicked to provide familiarity and uniformity.

Regarding dependent claims 6 and 7, Bharat et al. teach that *because the query topic Q can include a large number of terms, and because the "vocabulary" of the various pages can vary considerably, we prefer to use term frequency weighting. More specifically, we use cosine normalization in weighting both the query topic Q and the pages P because the deviation in term vector lengths is large, specifically: ... where* $w_{iq} = freq_{iq} \times IDF_i$, $w_{ij} = freq_{ij} \times IDF_i$, *$freq_{iq}$ is the frequency of (stemmed) term i in the query topic Q, $freq_{ij}$ is the frequency of term i in page j, and IDF_i is an estimate of the inverse document frequency (IDF) of the term i in the corpus of documents, for example, in our case, a large representative sample of Web pages* (Column 7, lines 10 – 29), which meet the limitation of **determining a first frequency at which the anchortext appears in the index; determining a second frequency at which each token derived from**

the anchortext appears in the index; and assigning a weight to the token, wherein the weight is a function of the first and second frequencies, and **dividing the first frequency by the second frequency to produce a weight quotient; and multiplying the weight quotient by an anchor text count for the token**.

Regarding dependent claim 8, Bharat et al. teach that *during a connectivity analysis phase, the remaining nodes of the pruned graph are then scored according to their connectivity to determine normalized hub and authority scores for the documents. The normalized scores are used to rank the documents* (Column 3, lines 31 – 35), which meet the limitation of **determining a normalized weight for each token**.

Regarding dependent claims 14 and 15, the claims incorporate substantially similar subject matter as claims 6 and 8, and are rejected along the same rationale.

Regarding independent claims 9 and 16, the claim incorporates substantially similar subject matter as claim 1, and is rejected along the same rationale.

(10) Response to Argument

Appellant argues that Bharat et al. do not teach **retrieving anchortext associated with at least one of the hyperlinks**, and **parsing the anchortext into one or more tokens** because Bharat et al. is silent regarding anchortext (p 6 - 8).

The Office disagrees.

First, it should be noted that Appellant does not cite where in the specification the alleged definition of anchortext appears. It is unclear how anchortext "fundamentally" excludes the URL itself.

It should be noted that a passage within the specification, p 2, lines 11 – 18, does not exclude anchortext from being the text of the actual URL. In contradistinction, the passage states that *in the page containing the link, usually there is some text associated with the link. In typical browsers the user clicks on this text to follow the link. This text is known as anchortext.*

In other words, anchortext is the text a user clicks to follow a link. Within the broadest, reasonable interpretation in light of the specification, the term anchortext is a term of art known to those of ordinary skill as evidenced by the Google definitions cited by the Office. Google shows that while anchortext can be non-URL text such as "linking text or anchor text" (p 1, under Web), it can also be the text of the URL such as "www.patrickgavin.com/SEO-Glossary.htm" (p 1, under first definition) in accordance with the first definition *Also known as Link Text, the clickable text of a hyperlink* (p 1). Furthermore, Google defines linking text or anchor text as simply *the text that is contained within a link* (p3). It should further be noted that all of the anchor text displayed in the bodies of the three pages are essentially URLs or the text thereof.

Therefore, the Office maintains that Bharat et al.'s teachings that *the score is a weighted sum of the number of directed edges to and from a node and the number of unique sub-strings of the URL that match a query term* (Column 3, lines 3 – 15), which meet the limitation of **retrieving anchortext** (*strings in the URL*) **associated with at**

least one of the hyperlinks (*directed edges to and from a node*), and **parsing the anchortext** (*string in the URL*) **into one or more tokens** (*sub-strings in the URL*).

Appellant incorrectly asserts that the Office meant to equate the claimed anchortext with the disclosed teaching of sub-strings in the URL (p 7, first full paragraph). The Office has sought to correct such confusion and reiterates that the string of the URL is equivalent to the anchortext and that the sub-strings of the URL are equivalent to the parsed tokens.

It should be noted that appellant now addresses the fact that the Office indeed interpreted and asserted so in its arguments that the term anchortext may encompass the text or strings of the URL (p 7, last paragraph). However, the appellant appears to be consumed with semantics.

The fact is Bharat et al. explicitly teaches taking a URL string, parsing the URL string into sub-strings and using those sub-strings as a weighted score to rank webapges. Thus, at the very least, it would have been obvious to a person of ordinary skill in the art to try parsing anchortext other than the strings of the URL in an attempt to provide an improved ranking system as a person with ordinary skill has good reason to pursue the kown options within his or her technical grasp. In turn, because anchortext as claimed has properties, simply a label or string, predicted by the prior art, it would have been obvious to parse all forms of anchortext.

Again, Appellant incorrectly asserts that the Office meant to equate the claimed anchortext with the disclosed teaching of sub-strings in the URL (p 8, last full

paragraph). The Office has sought to correct such confusion and reiterates that the string of the URL is equivalent to the anchortext and that the sub-strings of the URL are equivalent to the parsed tokens.

Appellant appears to argue that Bharat et al. does not teach that claimed indexing step because of the definition of anchortext and the appellant's mischaracterization of the Office's interpretation. It should be noted that the Appellant alleges that the Office relies on official notice to teach indexing in one breath (p 9, first full paragraph) but then admits that Bharat et al. indeed discloses a system the indexes documents in another breath (p 9 last paragraph).

Again, Appellant incorrectly asserts that the Office meant to equate the claimed anchortext with the disclosed teaching of sub-strings in the URL (p 9, last paragraph). The Office has sought to correct such confusion and reiterates that the string of the URL is equivalent to the anchortext and that the sub-strings of the URL are equivalent to the parsed tokens.

(11) Related Proceeding(s) Appendix

No decision rendered by a court or the Board is identified by the examiner in the Related Appeals and Interferences section of this examiner's answer.

For the above reasons, it is believed that the rejections should be sustained.

Respectfully submitted,

Nathan Hillery

Conferees:

/Doug Hutton/
Doug Hutton
Supervisory Primary Examiner
Technology Center 2100

Doug Hutton, Supervisory Patent Examiner, AU 2176

Stephen Hong, Supervisory Patent Examiner, AU 2178

STEPHEN HONG
SUPERVISORY PATENT EXAMINER